APPENDIX G

ANALYTICAL VARIABILITY IN REASONABLE POTENTIAL AND PERMIT LIMIT CALCULATIONS

This page intentionally left blank.

ANALYTICAL VARIABILITY IN REASONABLE POTENTIAL AND PERMIT LIMIT CALCULATIONS

Appendix G explains how analytical variability affects calculations used to determine reasonable potential and permit limits, and how such variability affects WET measurements. The appendix also considers suggested approaches to adjusting the reasonable potential and permit limit calculations to account for analytical variability. Only water quality-based effluent limitations are addressed because different considerations apply to technology-based limitations. While Appendix G addresses WET variability, its discussion and conclusions apply, with obvious modifications in terminology, to concentrations of chemical pollutants.

EPA has evaluated methodologies to adjust for analytical variability in setting permit limits. These methodologies would allow permit limits to exceed acute and chronic wasteload allocations (WLAs), sometimes two-fold or more. EPA believes that such approaches contradict the intent and practice of current guidance and regulations directed at preventing toxicity. The TSD calculations were carefully designed to avoid setting limits that allow a discharge to routinely exceed WLAs. Attempts to use an "adjusted," smaller estimate of variability in the first step of the effluent limit calculation (calculating the long-term average from the WLA) while using the variability of measured toxicity in the second step (calculating limits from the LTA), as done in the "adjustment approaches," will risk setting limits that exceed WLAs because the second variability factor is larger than the first. EPA also believes that the TSD statistical approach is adequately protective. On average, it achieves the desired level of protectiveness that is described in the NPDES regulations (40 CFR 122.44(d)) and EPA guidance.

This review did not evaluate the "conservativeness" of other components of WET limits, such as the acute-to-chronic ratio (ACR) for WET, the suggested WET criterion values (TUa = 0.3 and TUc = 1.0), and the methods of calculating the WLA using models of effluent dilution. Instead, this review took the WLAa (or WLAa,c) and WLAc as given and considered the TSD statistical method per se.

G.1 TSD Statistical Approach to Reasonable Potential And Limit Calculations

This appendix provides a simplified description of the TSD approach. That approach is more completely described in the *Technical Support Document for Water Quality-Based Toxics Control* (USEPA 1991a). Reasonable potential calculations are described in Section 3.3 of that document. The calculation is only one component of a reasonable potential determination. Permit limit calculations are described in Section 5.4 and Appendix E of the TSD.

To evaluate reasonable potential or calculate permit limits, one needs a coefficient of variation (CV) representing the variability of toxicity or a pollutant in the effluent discharge. The TSD recommends that the CV of measured effluent data be used in all reasonable potential and effluent limit calculations without attempting to "factor out" analytical variability. The specification of this CV is at issue in the alternatives to the TSD statistical procedures discussed later in this appendix.

G.1.1 Reasonable Potential

The goal of the TSD reasonable potential calculation is to estimate the probable value of an upper bound (e.g., 99th percentile) of toxicity in an effluent discharge using limited data. For whole effluent toxicity (WET), data are expressed in toxic units (TU) before calculating the CV. TU = (100/effect concentration). For chronic toxicity, TUc = 100/NOEC or 100/IC25. For acute toxicity, TUa = 100/LC50. The TSD calculations assume that effluent toxicity values follow a lognormal distribution, at least approximately. There is abundant evidence supporting the lognormal distribution, but the TSD also

acknowledges that other distributions might be found more appropriate if sufficient data can support the finding.

The sample CV of effluent monitoring data is obtained in TU. If there are fewer than ten data points, the TSD recommends a default CV of 0.6. The TSD recommends basing a calculated CV on at least ten data points, collected at the same time intervals as intended for monitoring.

Even if there are fewer than ten data points, the maximum value for the data (e.g., TU_{max}) is used to calculate a projected maximum value. A nonparametric, upper tolerance bound is calculated to infer the population percentile represented by TU_{max} with probability P: $X_{P,n} = (1 - P)^{1/n}$. For example, with probability 0.99 the largest of five observations will exceed the 39.8^{th} population percentile: $(1 - 0.99)^{1/5} = 0.398$. Next, the ratio between this percentile ($X_{P,n}$) and the population 99^{th} percentile is estimated using moment estimators for a lognormal distribution:

Reasonable potential multiplier = $X_{0.99} / X_P = \exp(Z_{99}\,\sigma - 0.5\sigma^2) / \exp(Z_P\,\sigma - 0.5\sigma^2)$.

Here, σ^2 is estimated as $\log(1 + CV^2)$, using the default CV if necessary. The maximum projected value is the product of the observed TU_{max} and the reasonable potential multiplier. This value may be compared to the WLA, which is based upon the criteria continuous concentrations (CCC) or criteria maximum concentration (CMC) and the appropriate dilution factors (if applicable). The projected maximum value also may be multiplied by a dilution factor and compared directly to the CMC or CCC (TSD Section 3.3, Box 3-2). The TSD recommends using TUa = 0.3 and TUc = 1.0 either as numeric toxicity criteria or as a means of interpreting the narrative "no toxics in toxic amounts" criteria.

G.1.1.1 Permit limit calculation

The first step in determining the appropriate water quality-based effluent limits for an effluent discharge is to calculate wasteload allocations WLAa and WLAc that correspond to the water quality criteria for acute exposures and chronic exposures or the ambient values used in interpreting narrative criteria (e.g., no discharge of toxic pollutants in toxic amounts). This step is distinct and separate from the "statistical" steps for calculating permit limits or reasonable potential. The WLAs are "givens" in the statistical calculations.

WLAa and WLAc are found through either a direct steady-state calculation or a dynamic model simulation. In either case, any applicable mixing zone and critical stream flows are taken into account. For WET, WLAa is converted to WLAa,c using an ACR. WLAs must not be exceeded if the water quality standards of the receiving water are to be met.

The essential idea behind setting a permit limit using the TSD approach is to find the lognormal distribution (i.e., its mean value or LTA) that would allow no more than a specified percentage of single observations to exceed the WLAa and no more than a specified percentage of the 4-day averages of observations to exceed the WLAc. If this percentage is set at 1 percent, for example, then the 99^{th} percentile of single observations must not exceed the WLAa, and the 99^{th} percentile of 4-day averages must not exceed the WLAc. The 4-day averaging period comes from the typical definitions of chronic exposure and the CCC. The CV has already indirectly specified the distribution's standard deviation. Together, the CV and the LTA specify the appropriate distribution completely.

The calculations which lead to finding the LTAa,c and LTAc (corresponding to the WLAa and WLAc) work in the following manner. The ratio between the LTA and a percentile (X_P) is called a variability factor (VF_P). The VF is calculated from the CV, the percentile (95^{th} or 99^{th}), and the averaging period [1 day (no averaging) or 4 days].

Thus, $LTA = X_P / VF_P$

If we set X_P equal to the WLAa, we find:

$$LTAa,c = WLAa / VF_{99,\ 1\text{-day}}$$
and
$$LTAc = WLAc / VF_{99,\ 4\text{-day}}$$

The smaller of the two LTAs is selected as the LTA used to calculate a limit. This step assures that the limits will exceed neither the WLAa nor the WLAc.

Having selected the smaller LTA, the VF calculation is reversed. Following the TSD recommendations,

$$\text{“Maximum Daily Limit” (“MDL”)} = LTA * VF_{99,\ 1\text{-day}}$$

and

$$\text{“Average Monthly Limit” (“AML”)} = LTA * VF_{95,\ N\text{-day}}$$
(based on N observations)

Note that in calculating the average limit the TSD recommends using a 95th percentile (rather than a 99th percentile) and the number of observations N for averaging may be less than four (although the TSD recommends $N \geq 4$ for purposes of calculating average limits). Limits calculated using the TSD-recommended approach are always equal to or less than the WLAa and WLAc.

G.1.1.2 Analytical variability in the TSD procedures

Analytical variability is a part of the variability of measurements used to analyze reasonable potential and set water quality-based limits. All components of variability that will enter into the permit development process are included in the measurements and calculations used to evaluate reasonable potential and set limits. This insures that the WLA is not exceeded.

Some laboratories have suggested alternative statistical calculations to EPA. Sections G.3 and G.4 discuss these approaches. These alternative calculations, however, would allow limits to exceed the WLA. When a sample effluent toxicity equals the WLA exactly, analytical variability would be expected to cause tests to exceed the WLA about half the time. Limits set above the WLA could allow routine exceedances of the WLA. In contrast, limits set using the TSD approach will provide some margin of safety between the limit and the WLA, guarding to some extent against analytical variability. On average, the TSD approach, employing the CV of measurements, is expected to ensure that the WLA is not exceeded when measured toxicities remain within the limits.

G.2 Background on Analytical Variability and Variability of Measurements

This section describes how analytical variability may cause the variance (σ^2) of measured values to exceed the variance of toxicity. This discussion will assume that WET tests for one discharge are conducted by one laboratory. Thus, “analytical variability” here will refer to within-laboratory variability (repeatability) of WET test results.

G.2.1 Components of Measurement Variability

The variance of monthly or quarterly measurements of effluent toxicity depends on at least two components: the variance of the toxicity, which changes over time, and the variance owed to the analytical process (including calibration, if applicable). One could also distinguish a third component—sampling variance—if simultaneous samples differ in toxicity. Herein, this component will not be examined separately, but is combined with the variance in toxicity over time.

A direct way to estimate the analytical component of variability is to analyze the same sample of effluent on different occasions so that the analytical method is the only source of measurement variance. The sample must be measured on different days because real samples are measured at intervals of weeks to months and the analytical process can change subtly over time. Unfortunately, effluent samples may not retain the same toxicity for long. Therefore, saving a batch of sample and analyzing it once a month for several months may over-estimate analytical variability. Analyzing two or three subsamples on the same date may underestimate analytical variability because the measurement system changes between sampling dates. The organisms, laboratory technicians and procedures, and laboratory materials may all change subtly over time. It would be reasonable to design a study that measures analytical variability in both ways, using effluent subsamples on one occasion and using the same (stored) effluent sample on separate occasions, attempting to bracket the correct value of analytical variance. EPA is not aware of any such studies. Reference toxicant samples are expected to have the same potency on different occasions and are used routinely for laboratory quality assurance of WET test methods. This document summarizes the variability resulting from repeated (usually monthly) WET testing of reference toxicant samples in the same laboratory.

G.2.2 Effect of Analytical Variability on Measured Values

Because of analytical variability the probability distribution of measured values Y is "wider" than the distribution of true values X. Thus, the mean and high percentiles of measurements will exceed the percentiles of the true values.

One component of the variance of measurements is analytical variance. Simple but plausible assumptions lead to the equation $V_Y = V_X + V_A$. In other words, the variance of a measurement Y (toxicity) is the sum of the variances for toxicity (V_X) and the analytical variance (V_A). When this equation is approximately correct, then one suitable estimate of V_X is ($V_Y - V_A$), where the parameters V_Y and V_A are replaced by their sample estimates. This estimate may be biased (i.e., inaccurate) to some degree. Similar reasoning about the mean (EY) leads to EY = EX. Then $V_Y = V_X + V_A$ can be divided by EX^2 to give $CV_Y^2 = CV_X^2 + CV_A^2$. This reasoning requires two assumptions: variance is constant and unrelated to the mean, and there is little or no correlation between X and the magnitude of the analytical error. When X is distributed lognormally, these assumptions are not true, but may be suitable for transformed values like log(Y) and log(X).

G.2.3 Analytical Variability and Self-monitoring Data

EPA determines compliance with a limit on the basis of self-monitoring data. No special allowance is made for analytical variability. This is accounted for by the TSD statistical procedures used to determine the need for limits and calculate permit limits.

The permittee must ensure that the toxicity in the discharge is never great enough to result in a compliance measurement that exceeds the permit limit. The maximum discharge toxicity allowed by the treatment system must incorporate a margin of safety to account for the sampling and analytical variability that attends compliance measurements. In other words, to avoid exceedances of a limit, a treatment system will be designed so that the maximum discharge toxicity is somewhat lower than the permit limit. Most industrial and municipal treatment facilities should be able to implement such a design. When they are not, appeals based on fundamentally different factors and economic hardships may be feasible.

G.2.4 Imprecision in WET Estimates, Reasonable Potential Determinations and Limits

Although WET tests provide protection against false positives, the estimates (NOEC, EC25, LC50) from WET tests, like all estimates based on limited data, are imprecise. That is, the exact level of toxicity in a sample is estimated with "error" (imprecision). This imprecision can be reduced by providing a suitable number of organisms and replicates for each test. The numbers required for EPA WET method test

acceptability are *minimums*. Test precision will be approximately proportional to the square root of the number of replicates. Thus, a doubling of replication may increase the precision of a test endpoint response (survival, growth, reproduction) to roughly 70 percent of its former level. For example, consider these calculations for fathead minnow growth (USEPA 1994a, pp. 102-105): the standard error of the difference between a treatment and the control is $Sw\sqrt{(1/n_T + 1/n_C)}$, which in one test took the value $(0.0972)\sqrt{(1/4 + 1/4)} = (0.0972)(0.707) = 0.0687$. If the root mean squared error Sw had been the same but the number of replicates had been doubled, the standard error would have been 0.0486. Dunnett's critical value would have been 2.24 instead of 2.36, and the MSD 0.109 instead of 0.162. With a doubling of replication, the test would be able to detect a 16-percent reduction from the control rather than a 24-percent reduction.

For reasonable potential and limit calculations, WET data are accumulated over a year or more to characterize effluent variability over time. This sampling program may not fully characterize effluent variability if too few samples are taken, if the sampling times and dates are not representative, or if the duration of the sampling program is not long enough to represent the full range of effluent variability. For reasonable potential and limits, the key quantity being estimated is the variance (or CV). A large number of samples is required to estimate a variance or CV with much precision. Confidence intervals for the variance and CV can be calculated easily and carried through the calculations for reasonable potential and effluent limits (Section G.1). Even when assumptions are not strictly met, this information may provide a useful perspective on the uncertainty of the calculation.

G.2.5 Between-laboratory Variability in Reasonable Potential and Permit Limit Calculations

It is inappropriate to use estimates of between-laboratory variability in calculations of reasonable potential and permit limits. Such estimates do not represent the variability affecting measurements of effluent discharge toxicity. In most cases, only one laboratory will produce the data for one discharge. In some cases, there will be a change of laboratory over time, which needs to be handled case-by-case. Using estimates of between-laboratory variability to represent the analytical component of variance for one discharge is equivalent to assuming that each new sample is sent to a new laboratory selected at random from the population of laboratories conducting the test method. This approach does not occur in practice.

Between-laboratory differences in test sensitivity are important and need to be addressed. To some extent, apparent differences in sensitivity between laboratories (Warren-Hicks et al. 1999) may be owed to several factors, including use of unstable reference toxicants like SDS (Environment Canada 1990), errors in calculating and recording stock concentrations (Chapter 3 of the Variability Guidance, SCTAG 1996), differences in dissociation and bioavailability of metal ions, comparisons of non-comparable ionic forms (e.g., potassium chromate versus potassium dichromate, SCTAG 1996), use of different waters, health of organisms, and varying techniques.

Within-laboratory variability should be reflected in regulatory calculations. If the data being used for reasonable potential or permit limit calculations consist of effluent measurement data reported by two or more laboratories, there are ways to account for between-laboratory differences:

- If the same laboratories are used in the same proportion or frequency, and the measurements for different laboratories represent different sampling dates, the measurement data may be treated as if they come from one laboratory. This may increase the estimated variance and the average monthly limit, which is not in the interest of the permittee. It would be better to select one laboratory, based on the variance of its reported reference toxicant test results.

- If only one laboratory has reported data on each date, with the different laboratories either reporting over different time spans or over the same time span on alternate dates, EPA recommends a pooled

estimate of variance. Calculate the sample variance S^2 for log(TU) separately for each laboratory, and combine the data in the following formula:

$$\text{pooled variance of } \log(X) = [(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2] / [(N_1 - 1) + (N_2 - 1)]$$

(i.e., the analogous formula for more than two laboratories). The same result can be obtained by conducting a one-way analysis of variance on log(X) and using the mean squared error. This approach would be undesirable if the different laboratories sampled times or time spans that were known or expected to differ in the average or variance of TU. In that case, one would pool the data, treating it as if it had come from one laboratory (see above).

A change of testing laboratory by a permittee may result in a change in analytical (within-laboratory) variability of measurements and a change in "sensitivity." The average effect concentration may change. There may be between-laboratory differences in sensitivity to some toxicants, such as metals (Warren-Hicks et al. 1999).

Ideally, a permittee will anticipate a change of the testing laboratory. Permittees should compare reference toxicant test data from current and candidate replacement laboratories, selecting a laboratory with acceptable variability and a similar average effect concentration. Regulatory authorities should compare reference toxicant data for old and new laboratories when interpreting a series of WET test results that involves a change of laboratory.

Some areas may help reduce laboratory differences in average effect concentration for the same reference toxicant test protocol. These include standardization and reporting of stock culture conditions (such as loading, age structure, age-specific weight, and other conditions), standardization of dilution water for reference toxicant tests, and reporting to verify such practices. Other areas for consideration include test protocols, test acceptability criteria, and dilution water. Another approach that could be evaluated further is conducting a reference toxicant test with each effluent test, and normalizing the effluent response using the toxicant response.

G.3 Adjustment Approaches To Account For Analytical Variability in Setting Permit Limits

G.3.1 Adjustment Approaches To Account for Analytical Variability

Methods have been proposed for determining reasonable potential and calculating permit limits by adjusting the calculations based on analytical variability. The more general principles are discussed here, details of these methods are outlined in Section G.4. The focus of these discussions is the limit calculation, although similar principles apply to the reasonable potential calculation.

The idea behind the proposed "adjustment methods" for calculating water quality-based effluent limits is to estimate the distribution of toxicity values using data on measured effects concentrations and analytical variability, and then to factor out analytical variability from some steps in the process of calculating limits. In proposed adjustment methods for calculating effluent limitations one would (1) estimate the variance of effluent concentrations (this entails subtracting an estimate of the analytical variance from the variance of effluent measurements, e.g., $V_X = V_Y - V_A$, or an equivalent calculation using CVs); (2) calculate the LTAa and LTAc using the TSD approach and the adjusted variance V_X; and (3) calculate the limit (from the lower of the two LTAs) using the variance of measurements V_Y. Because the V_Y necessarily exceeds V_X, these methods would result in limits that would exceed calculated WLAs, depending on other assumptions made in the limit calculations. As a result, the discharge may allow instream WET to routinely exceed the criterion limits, a condition that should not occur.

G.3.2 Adjustment Equations

As noted above, the adjustment approaches are based on the TSD statistical approach, modified to subtract analytical variability from the LTA calculation. These approaches refer to V_X as the "true" variance. In what follows, the sample estimate of V_X is S^2_{True}. Thus, $S^2_{True} = S^2_{Meas} - S^2_{Analy}$ (where S^2 is the sample estimate of variance) is used to calculate the LTAs and S^2_{Meas} is used to calculate the limits from the smallest of the two LTAs. The TSD equations as applied to WET would be adjusted as follows:

When the LTAa,c is the smallest LTA,

$$MDL = WLAa,c * (VF_{99,\ 1\text{-}day,\ Meas} / VF_{99,\ 1\text{-}day,\ True})$$
$$AML = WLAa,c * (VF_{95,\ N\text{-}day,\ Meas} / VF_{99,\ 1\text{-}day,\ True})$$

When LTAc is the smallest LTA (and assuming that the chronic criterion is a 4-day average)

$$MDL = WLAc * (VF_{99,\ 1\text{-}day,\ Meas} / VF_{99,\ 4\text{-}day,\ True})$$
$$AML = WLAc * (VF_{95,\ N\text{-}day,\ Meas} / VF_{99,\ 4\text{-}day,\ True})$$
$$\text{where } N = \text{samples/month (for purposes of AML calculation)}$$

The VF (variance factor) is the ratio of a percentile to a mean, in this case for the lognormal distribution.

$$VF_{99,\ 1\text{-}day,\ Meas} = \exp(Z_{99} S_{Meas} - 0.5S^2_{Meas})$$
$$VF_{99,\ 1\text{-}day,\ True} = \exp(Z_{99} S_{True} - 0.5S^2_{True})$$
$$VF_{95,\ n\text{-}day,\ Meas} = \exp(Z_{95} S_{n\text{-}day,\ Meas} - 0.5S^2_{n\text{-}day,\ Meas})$$
$$VF_{99,\ 4\text{-}day,\ True} = \exp(Z_{99} S_{4\text{-}day,\ True} - 0.5S^2_{4\text{-}day,\ True})$$

while
$$S^2_{Meas} = \log(1 + CV^2_{Meas})$$
$$S^2_{True} = \log(1 + CV^2_{True})$$
$$S^2_{N\text{-}day,\ Meas} = \log(1 + CV^2_{Meas}/N)$$
or
$$S^2_{N\text{-}day,\ Meas} = S^2_{Meas}/N = \log(1 + CV^2_{Meas})/N$$
$$S^2_{4\text{-}day,\ True} = \log(1 + CV^2_{True}/4)$$
or
$$S^2_{4\text{-}day,\ True} = S^2_{True}/4 = \log(1 + CV^2_{True})/4$$

G.3.3 Consequences of Adjustment Approaches

As an example of the consequences of applying an adjustment methodology to water quality-based effluent limit calculations, one may consider the following scenario. In this scenario, such a methodology would allow calculation of an average monthly limit (AML) exceeding the chronic WLA (a four-day average value) even when sampling frequency for the calculation is set at the recommended minimum of four samples per month. It is acceptable for the MDL (a single sample) to exceed the chronic WLA or for the AML to exceed the chronic WLA if the AML calculation is based on less than four samples per month. Note, however, that the TSD recommends always assuming at least four samples per month when calculating the AML.

Table G-1 below offers an example of MDLs and AMLs calculated using the TSD approach and an approach that adjusts the CV for analytic variability. This adjustment would allow effluent limits that exceed the WLA on the premise that analytical variability tends to make measured values larger than actual effluent values. Thus, this approach assumes that the "true" monthly average would be below the WLAc even though the limit and the measured monthly average may be above the WLAc.

EPA believes that these assumptions are invalid. Therefore, EPA cannot recommend an approach that makes such assumptions as part of national guidance to regulatory authorities. EPA is not recommending national application of an "adjustment approach" to either reasonable potential or effluent limit calculation

procedures. EPA continues to recommend the TSD approach, which ensures that effluent limits and, thereby, *measured* effluent toxicity, are consistent with calculated WLAs.

Table G-1. Sample Effluent Limit Calculations Using EPA's TSD Approach and an Adjustment Approach (USEPA 1991a)

WLA_c	Probability Basis	Approach	LTA_c	MDL	AML
10	MDL = 99th percentile AML = 95th percentile	TSD	4.4	17.6	7.7
10	MDL = 99th percentile AML = 95th percentile	Adjustment approach	6.43	25.8	***11.2*** *
10	MDL = 99th percentile AML = 99th percentile	TSD	4.4	17.6	9.99
10	MDL = 99th percentile AML = 99th percentile	Adjustment approach	6.43	25.8	***14.6*** *

Assumptions: Chronic LTA/WLA controls calculations, WLA = 99th percentile probability basis, n = 4 (sampling frequency for AML calculation), Total CV = 0.8 and Adjusted CV = 0.4 are used in calculations.
(*) These numbers exceed the WLAc.

G.3.4 Related Concerns

In addition to addressing the differences between measured and "true" values in the reasonable potential and effluent limit calculations, related concerns regarding WET testing and the water quality-based effluent permits process have been raised as reasons for adjusting the TSD statistical procedures.

G.3.4.1 Compounding protective assumptions

Approaches to "account for analytical variability" by adjusting the calculations for reasonable potential and limits usually state that several conservative assumptions are employed. In the TSD approach, a water quality-based effluent limit is the result of three key components: (1) a criterion concentration; (2) a calculated dilution or mixing-zone factor; and (3) a statistical calculation procedure that employs a CV based on effluent data. The conservative assumptions cited may involve deriving the criterion concentration, and assuming dilution and low-flow conditions, in addition to the probability levels used in the TSD statistical calculations. Even if these assumptions were considered conservative, the TSD statistical procedure remains valid. As explained above, the TSD statistical approach is *appropriately* protective, provided that the WLA is accepted as given. It is inappropriate for regulatory authorities to modify the TSD's correctly conceived statistical approach in order to compensate for assumptions intrinsic to derivation of the WLA that are perceived as over protective. Therefore, EPA does not believe that it is appropriate to adjust the TSD statistical methodology for conducting reasonable potential and calculating permit limits to address concerns about how WLAs are calculated.

G.3.4.2 Test sensitivity and method detection limit

EPA does not employ method detection limits (MDLs: 40 CFR part 136 Appendix B) for WET methods. For effect concentrations derived by a hypothesis test (LOEC and NOEC), the alpha level of the test provides one means of providing a functional equivalent of an MDL. The hypothesis test prescribed in the method provides a high level of protection from "false positives." For point estimates (ECp, ICp, LCp), a valid confidence region provides the equivalent of a hypothesis test. EPA will provide clarification regarding when confidence intervals are not or cannot be generated for point estimation procedures, including the ICp procedure. This variability guidance cites recommendations (Chapman et al. 1996a, Baird et al. 1996, Bailer et al. 2000) regarding alternative point estimation methodologies.

While protecting against false positives, hypothesis tests and confidence intervals, will provide little protection from toxicity unless the test method is designed to detect a suitable effect size. The two most commonly used chronic tests are incapable of routinely detecting effects of 20 percent to 30 percent (Denton and Norberg-King 1996) when employed by many laboratories using the minimum recommended number of replicates and treatments. To provide suitable test sensitivity, regulatory authorities should consider requiring more replication, a suitable minimum significant difference (MSD), or suitable effect sizes and power, particularly for the control and IWC test concentrations (e.g., Denton and Norberg-King 1996; Washington State Department of Ecology 1997, Ch. 173-205 WAC). It may be desirable to specify that a statistically significant effect at the IWC must exceed some percentage difference from the control before it is deemed to have regulatory significance. Combining these approaches, an effective strategy would require that a test consistently be able to detect the smallest effect size (percent difference between the control and the IWC) that would compromise aquatic life protection, and to disregard very small, statistically significant effects. To further these ends, this guidance document sets an upper limit to the value of MSD/(Control Mean), defining the maximum acceptable value. This document also sets a lower limit to the effect size, defined by 100×(Control Mean - Treatment Mean)/(Control Mean), which can be regarded as "toxic" in a practical sense (see Section 6.4).

The alpha level of a hypothesis test or confidence interval cannot be decreased from that level (α = 0.05) recommended for WET methods without sacrificing test power and sensitivity of the method. Alpha should not be decreased without a corresponding increase in sample size that would preserve the power to detect biologically significant effects. EPA will issue guidance on when the nominal error rate (alpha level) may be adjusted in the hypothesis test for some promulgated WET methods (USEPA 2000a).

G.4 Technical Notes on Methods of Adjusting For Analytical Variability

This section describes and comments on several adjustment methodologies suggested to EPA as alternatives to the TSD statistical calculations.

G.4.1 Notation

Explanations may help clarify the notations in this section. The symbols VX, V[X], and σ^2_X all mean: the variance of X. Standard deviation (σ_X) is the square root of the variance. The mean (average) is symbolized as EX and also as μ_X.

When X is lognormally distributed, there is a potential for confusing the mean and variance of log(X) with the mean and variance of X. Typically (and in the TSD), when X is lognormally distributed, the parameters will be given for log(X) as follows: X ~ lnorm(μ, σ). This is read as "X is distributed lognormally with the mean of log X equal to μ (mu) and the standard deviation of log X equal to σ (sigma)." Better notation would be X ~ lnorm(μ_{logX}, σ_{logX}); recommended terms for the parameters are "mu-logX" and "sigma-logX." The mean and variance of X for this distribution are

$$\mu_X = EX = \exp(\mu_{logX} + 0.5*\sigma^2_{logX})$$
$$\sigma^2_X = VX = \exp(2*\mu_{logX} + \sigma^2_{logX}) * [\exp(\sigma^2_{logX}) - 1]$$

To avoid confusion, the symbols EX and VX are used in preference to μ_X and σ^2_X to signify the mean and variance of X. Usually, mu and sigma are used only as symbols for the mean and standard deviation of log(X), that is, μ_{logX} and σ_{logX}, in the context of lognormal distributions. Below, μ_{logX} and σ_{logX} are abbreviated to μ and σ, with the addition of subscripts like "Effl" and "Meas" to further distinguish the intended quantity.

CV may be used to symbolize parametric values or their sample estimates, with the meaning indicated in the text. Symbols S^2_{Effl} , S^2_{Meas} , and S^2_{Analy} will represent sample estimates of variances $\sigma^2_{logX, Effl}$, $\sigma^2_{logX, Meas}$, and $\sigma^2_{logX, Analy}$.

G.4.2 General Comments on Analytical Variance as a Component of the Variance of Measurements

Two simple models lead to the same equation. The first model assumes that each measurement Y is the sum of a concentration X and an analytical error ϵ, that is $Y = X + \epsilon$. The analytical error ϵ may be positive or negative and has mean zero and variance V_A. X and ϵ are uncorrelated. (This is a strong assumption; it may be approximately correct only for some transformation of the data.) Then $V_Y = V_X + V_A$. The second, hierarchical, model assumes that X follows a distribution P_X with mean and variance E_X and V_X. Each measurement Yt (t indexes the time of measurement) follows another distribution having mean Xt and variance V_A. V_A is assumed to be constant, independent of Xt. (This is a strong assumption which may be approximately correct only for some transformation of the data.) Then, it can be shown that $V_Y = V_X + V_A$. The same models and assumptions lead to EY = EX. These models and assumptions are not correct when X is lognormally distributed. In that case, the models might provide reasonable approximations to the behavior of log(X) and log(Y). If EY = EX and $V_Y = V_X + V_A$ are both correct, then $V_Y = V_X + V_A$ can be divided by EX^2 to give $CV_Y^2 = CV_X^2 + CV_A^2$. In this case, the parameters V_X and CV_X^2 might be estimated by using sample estimates in the expressions $(V_Y - V_A)$ and $(CV_X^2 - CV_A^2)$, respectively. Such estimates will be somewhat biased.

G.4.3 Commonwealth of Virginia Approach

The Commonwealth of Virginia Toxics Management Program Implementation Guidance (1993) (revised on August 25, 1994) prescribes a method of accounting for analytical variability of WET data. A synopsis of the method follows. Symbolic notation has been changed; the numbered "steps" below were created for this synopsis.

1. Obtain the CV of WET monitoring data. This will be 0.6 (default value) if fewer than ten data are available. If there are at least ten data, a computer program (*described in Guidance Memo 93-015*) is used. *"Only acute test data are considered here because the LC_{50} is a statistically derived point estimate from a continuous data set. Also, the LC_{50}s must be real numbers. Values reported as '> 100%' should not be used in the calculation. .... Enter either LC_{50}s or TU_as for the most sensitive species into the program."* [Comments on Step 1: LC50 and TU values are not equivalent; they will not have the same CV values. The exclusion of ">100%" values will tend to bias the CV of TUs toward larger values.]

2. Calculate $S^2_{logX,\ Effl} = S^2_{logX,\ Meas} + S^2_{logX,\ Analy}$, using $S^2_{logX,\ Analy} = 0.20$. If $CV_{X,\ Meas} < 0.47$ (implying that $S^2_{logX,\ Meas} < 0.20 = S^2_{logX,\ Analy}$), instead use $S^2_{logX,\ Effl} = S^2_{logX,\ Meas}$. (These subscripts are not used in the Guide.) The value for $S^2_{logX,\ Analy}$ is based on data provided by several laboratories conducting tests for Virginia permits for the five most common species, using cadmium chloride as the reference toxicant. The Guide states that these data yielded a geometric mean CV_X of 0.47, and $0.20 = \ln(1 + 0.47^2)$; the last formula is the relation between the parametic variance and CV of a lognormal variate. [Comments on Step 2: The calculations should employ sample variances of log(TU), not sample CVs, in the interest of accuracy and precision. The estimate $S^2_{logX,Effl}$ is a discontinuous function, decreasing toward zero as $S^2_{logX,\ Meas}$ decreases toward 0.2, then jumping to 0.2 and decreasing again toward zero as $S^2_{logX,\ Meas}$ decreases further. The default value of $S^2_{logX,\ Effl}$ becomes $\ln(1 + 0.60^2) - \ln(1 + 0.47^2) = 0.11$.]

3. Calculate LTAa,c and LTAc as in the TSD, using $S^2_{logX,\ Effl}$ instead of $S^2_{logX,\ Meas}$, and using Z_{97}, the 97th percentile Z-statistic, instead of Z_{99}. WLA and LTA values are in units of TUc. The smaller of LTAa,c and LTAc is selected as LTA_{min}.

4. Calculate the "MDL" limit from LTA_{min} as in the TSD, now using $S^2_{logX,\ Meas}$ rather than $S^2_{logX,\ Effl}$ and still using the 97th percentile Z-statistic. No procedure is described for a limit of averages ("AML").

   By using this procedure, the WLAa,c may be exceeded when the CV of measurements exceeds 0.47 (because then the estimate $S^2_{logX,\ Effl} < S^2_{logX,\ Meas}$). The maximum ratio of Limit to WLA occurs when the CV of observations is just over 0.47, when the ratio of Limit to WLA is just over 2. Numerical evaluations (Table G-2) show that the daily limit can exceed the WLAa,c. The daily limit (DL or MDL) should be compared to the WLAa,c. It is not unusual for the daily limit to exceed the WLAc when LTAc is smaller than LTAa,c. This outcome does not necessarily indicate a problem. Instead, the regulatory authority should compare the average limit to WLAc in this case (see "Modified TSD Approach" below).

Table G-2. Numerical Effect of State of Virginia WET Limit Calculation on Ratio of Daily Limit to WLA

CV_{Meas}	S^2_{Effl}	$S^2_{Effl,\ 4\text{-day average}}$	Ratio of Daily Limit to WLAa,c	Ratio of Daily Limit to WLAc
0.10	0.01	0.00	1.00	1.09
0.20	0.04	0.01	1.00	1.19
0.30	0.09	0.02	1.00	1.27
0.40	0.15	0.04	1.00	1.35
0.45	0.18	0.05	1.00	1.38
0.470	0.1996	0.0538	2.097	1.393
0.471	0.0004	0.0002	2.026	2.042
0.50	0.02	0.01	1.65	1.87
0.60	0.11	0.03	1.39	1.76
0.70	0.20	0.06	1.28	1.74
0.80	0.29	0.09	1.22	1.72
0.90	0.39	0.13	1.18	1.71
1.00	0.49	0.17	1.16	1.70

The State of Virginia Guide, Appendix D, also states: "*Because the statistical approach evaluates both acute and chronic toxicity of the effluent, only one limit is necessary to protect from both acute and chronic toxicity. The limit is expressed only as a maximum daily limit (MDL) because the frequency of monitoring will typically be less than once per month. If the testing is to be monthly, then the MDL can also be expressed as an average monthly limit (AML).*" [Comment: a single MDL limit is not as protective as the combination of limits, one for single observations (MDL) and another for averages (for example, the quarterly or annual average). Refer to the TSD (USEPA 1991a, Section 5.3).]

G.4.4 Rice Approach

James K. Rice's unpublished draft, "Laboratory QC and the Regulatory Environment: Relation Between Method Performance and Compliance" prescribes a method of accounting for analytical variability of WET data. The document was provided with a notation that the typescript was originally submitted to EPA as a comment on the draft "TSD," presumably in the period 1989 to 1991. A synopsis of the method follows. The numbered "steps" below were created for this synopsis. Calculations and symbols have been

simplified. This synopsis omits many detailed observations that provide context and guidelines for readers intending to apply Rice's method.

1. Obtain the CV of WET monitoring data (measured values), and the CV of the analytical method, in symbols $CV_{X, Meas}$ and $CV_{X, Analy}$. Sample size is not addressed, but the text indicates that "a large number" of measurements are needed to characterize variability and bias.

2. Solve for $CV_{X, Effl}^2$ in $CV_{X, Meas}^2 = CV_{X, Analy}^2 + CV_{X, Ttue}^2 + (CV_{X, Analy}^2 * CV_{X, Effl}^2)$, after substituting the sample estimates of $CV_{X, Meas}^2$ and $CV_{X, Analy}^2$. Thus, solve

   $$CV_{X, Effl}^2 = (CV_{X, Meas}^2 - CV_{X, Analy}^2) / (1 + CV_{X, Analy}^2).$$

   [Comment: This formula assumes a model such as Measurement = (Concentration * Recovery), with multiplicative errors for Concentration and Recovery. This is one plausible model, especially for data that are distributed lognormally. Another plausible model would lead to the formula $CV_{X, Meas}^2 = CV_{X, Analy}^2 + CV_{X, Ttue}^2$.]

3. Calculate LTA values as in the TSD, using $CV_{X, Effl}$ instead of $CV_{X, Meas}$, and use Z_{99}, the 99th percentile Z-statistic. First calculate $\sigma^2_{logX, Effl} = \ln(1 + CV_{X, Effl}^2)$ for the variance of log(TU), and $\sigma^2_{logX, Effl, n} = \ln(1 + (CV_{X, Effl}^2)/n)$ for an n-day average. Then $LTA_{Effl} = WLA * \exp(0.5\sigma^2_{logX, Effl, n} - Z_P \sigma_{logX, Effl, n})$. Rice then calculates $LTA_{meas} = (R/100) * LTA_{Effl}$, where R is the percent recovery of the analytical method. [Comments: Many chemical methods are now calibrated instrumentally so that E[R] = 100 percent. It will be assumed herein that R = 100 percent for WET methods. There is no discussion of, or accounting for, the sampling error (the uncertainty) that attends the estimates of R or σ^2, of the sample sizes required to estimate these well. The example does not encompass the derivation and comparison of acute versus chronic LTAs using estimates of the variance of single observations and averages and selection of the smaller one, as in the 1991 TSD. Rice's method could easily be modified for the current TSD approach (see for example, the State of Virginia method, above).

4. Calculate the MDL and AML limits from the LTA as in the TSD, now using $\sigma^2_{logX, Meas}$ rather than $\sigma^2_{logX, Effl}$, and using the 99th percentile Z-statistic. Thus,

   $$MDL = LTA_{meas} * \exp(-0.5\sigma^2_{logX, Meas, 1} + Z_P \sigma_{logX, Meas, 1})$$
   $$AML_n = LTA_{meas} * \exp(-0.5\sigma^2_{logX, Meas, n} + Z_P \sigma_{logX, Meas, n})$$

   Using this procedure, the limits exceed the WLAc.

   $$MDL = WLAc * (VF_{.99, 1, Meas} / VF_{.99, 4, Effl}) > WLAc$$
   $$AML_n = WLAc * (VF_{.99, n, Meas} / VF_{.99, 4, Effl}) > WLAc \text{ if } n \leq 4$$

   The AML can exceed WLAc even if n >4, depending upon the variance values. Because the current TSD approach of comparing LTAa,c and the LTAc had not been developed by the time of Rice's report, he did not apply his procedure to the WLAa,c.

G.4.5 Amelia River Report

The Amelia River Report (USEPA 1987, Appendix G) describes a similar approach, estimating $S^2_{logX, Effl} = S^2_{logX, Meas} + S^2_{logX, Analy}$ (without any provision for the case $S^2_{logX, Meas} \leq S^2_{logX, Analy}$), calculating LTA from WLA using $S^2_{logX, Effl}$, and calculating the limits using $S^2_{logX, Meas}$.

G.4.5.1 Modified TSD approach

The methods described above predate the current TSD statistical approach and differ from it. As noted in the previous section, one could consider how the current TSD statistical approach could be modified to account for analytical variability using the same principles. The LTAs would be calculated using a variance estimate $S^2_{Effl} = S^2_{Meas} - S^2_{Analy}$, the smallest would be selected, and limits would be calculated from the smaller LTA using S^2_{Meas}. Table G-3 compares the current and modified calculations for whole effluent toxicity. Numerical calculations appear in Tables G-4 and G-5.

Table G-3. A Comparison of the Current TSD Calculation of Limits with a Modification That Takes into Account the Analytical Variability

Method	Smallest LTA	Limits
TSD statistical approach	LTAa,c	MDL = WLAa,c ($VF_{.99, 1, Meas}$ / $VF_{.99, 1, Meas}$) = WLAa,c AML = WLAa,c ($VF_{.95, N, Meas}$ / $VF_{.99, 1, Meas}$) < WLAa,c
	LTAc	MDL = WLAc ($VF_{.99, 1, Meas}$ / $VF_{.99, 4, Meas}$) < or > WLAa,c AML = WLAc ($VF_{.95, N, Meas}$ / $VF_{.99, 4, Meas}$) < WLAc
TSD modified to use S^2_{Effl} to calculate LTA	LTAa,c	MDL = WLAa,c ($VF_{.99, 1, Meas}$ / $VF_{.99, 1, Effl}$) > WLAa,c AML = WLAa,c ($VF_{.95, N, Meas}$ / $VF_{.99, 1, Effl}$) < or > WLAa,c
	LTAc	MDL = WLAc ($VF_{.99, 1, Meas}$ / $VF_{.99, 4, Effl}$) < WLAc AML = WLAc ($VF_{.95, N, Meas}$ / $VF_{.99, 4, Effl}$) < or > WLAc

Symbols for estimates based on data (sample estimates):

S^2_{Meas} sample variance of natural logs of measured TUs
S^2_{Analy} sample variance of natural logs of measurements on the same or TU
S^2_{Effl} estimate of variance of natural logs of TUs
$S^2_{Effl} = S^2_{Meas} - S^2_{Analy}$

$VF_{P, N, xxxx} = \exp(Z_P S_{xxx, N} - 0.5 S^2_{xxx, N})$ estimates the ratio of the P-th percentile to the mean for a lognormal variate: the P-th percentile is $\exp(\mu + Z_P \sigma)$ and the mean is $\exp(\mu + 0.5\sigma^2)$. The mean of a 4-day average of lognormal observations is assumed to be lognormal (Kahn, H.D., and M.B. Rubin. 1989. Use of statistical methods in industrial water pollution control regulations in the United States. *Environmental Monitoring and Assessment* 12:129-148).

The variance estimates may change with and be a function of the TU.
"N" is the number of samples (measurements) intended for use in determining compliance with the average limit, not the number of data used to calculate the sample variances used in setting limits.
It can be shown that LTAc < LTAa,c implies that WLAc < WLAa,c

For WET, WLAa,c = WLAa * ACR. It is assumed that the variance of observations (S^2_{Meas}) equals or exceeds the analytical variance (S^2_{Analy}). Numerical comparisons appear in Tables G-2 to G-4.

Calculations in Tables G-4 and G-5 show the numerical effect of adjustment on permit limits in relation to the WLA. These tables show the ratio of the limit to the WLA. For these calculations, S^2_{Meas} was calculated as $\log(1 + CV^2_{Meas})$, while $S^2_{Meas, 4\text{-}day} = \log(1 + CV^2_{Meas}/4)$, giving slightly different numerical results than if $S^2_{Meas, 4\text{-}day} = S^2_{Meas}/4 = \log(1 + CV^2_{Meas})/4$. The first formula is prescribed in the TSD, Box 5-2 and Table 5-1. The tables show the combinations of CV values used for CV_{Meas} and CV_{Analy}. The variance of TUs was calculated as $S^2_{Effl} = S^2_{Meas} - S^2_{Analy}$ using $S^2_{Meas} = \log(1 + CV^2_{Meas})$ and $S^2_{Analy} = \log(1 + CV^2_{Analy})$.

Table G-4. Ratio of MDL to WLA-LTA from WLA and CV_{Effl} and Limit from LTA and CV_{meas}

	LTAac is Smallest Ratio is MDL:WLAa,c					**LTAc is Smallest Ratio is MDL:WLAc**				
	CV_{Analy}					**CV_{Analy}**				
CV_{Meas}	0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5
0.1	1.25	0.00	0.00	0.00	0.00	1.25	0.00	0.00	0.00	0.00
0.2	1.06	1.55	0.00	0.00	0.00	1.28	1.55	0.00	0.00	0.00
0.3	1.04	1.17	1.90	0.00	0.00	1.38	1.47	1.90	0.00	0.00
0.4	1.03	1.11	1.31	2.28	0.00	1.48	1.55	1.69	2.28	0.00
0.5	1.02	1.09	1.22	1.48	2.68	1.58	1.63	1.73	1.93	2.68
0.6	1.02	1.07	1.16	1.33	1.65	1.66	1.70	1.79	1.93	2.18
0.7	1.01	1.06	1.13	1.26	1.47	1.72	1.76	1.83	1.94	2.12
0.8	1.01	1.05	1.11	1.21	1.37	1.77	1.81	1.87	1.96	2.10
0.9	1.01	1.04	1.10	1.18	1.30	1.81	1.84	1.90	1.98	2.09
1.0	1.01	1.04	1.08	1.16	1.26	1.84	1.86	1.91	1.98	2.08

[a] The LTA was calculated using the WLA and $_{Cveffl}$. The limit was calculated using the LTA and $CV_{meas.}$

Table G-5. Ratio of AML to WLA

	LTAa,c is smallest ratio is AML:WLAa,c					**LTAc is smallest ratio is AML:WLAc**				
	CV_{Analy}					CV_{Analy}				
CV_{Meas}	0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5
0.1	1.08	0.00	0.00	0.00	0.00	1.08	0.00	0.00	0.00	0.00
0.2	0.80	1.17	0.00	0.00	0.00	0.96	1.17	0.00	0.00	0.00
0.3	0.69	0.78	1.26	0.00	0.00	0.92	0.98	1.26	0.00	0.00
0.4	0.61	0.66	0.78	1.36	0.00	0.89	0.93	1.01	1.36	0.00
0.5	0.55	0.59	0.66	0.80	1.45	0.85	0.88	0.94	1.05	1.45
0.6	0.51	0.53	0.58	0.66	0.82	0.83	0.85	0.89	0.96	1.08
0.7	0.47	0.49	0.53	0.58	0.68	0.80	0.82	0.85	0.90	0.98
0.8	0.44	0.46	0.49	0.53	0.60	0.77	0.79	0.82	0.86	0.92
0.9	0.42	0.43	0.45	0.49	0.54	0.75	0.76	0.79	0.82	0.87
1.0	0.40	0.41	0.43	0.46	0.50	0.73	0.74	0.76	0.79	0.83

NOTE: If the AML were set at a 99th percentile value, all ratios would exceed 1.00. It is not surprising that the ratio in the table for AML is less than 1, should not come close to one, because the 95th percentile was used in the second part of the equation. The ratio should be constantly less than one in order to protect water quality criteria.

[a] The LTA was calculated using the WLA and $_{Cveffl}$. The limit was calculated using the LTA and $CV_{meas.}$